

Exploring Spatial Patterns and Topological Significance in Codon Group Distribution: A Multidimensional Approach

Douglas C. Youvan

doug@youvan.com

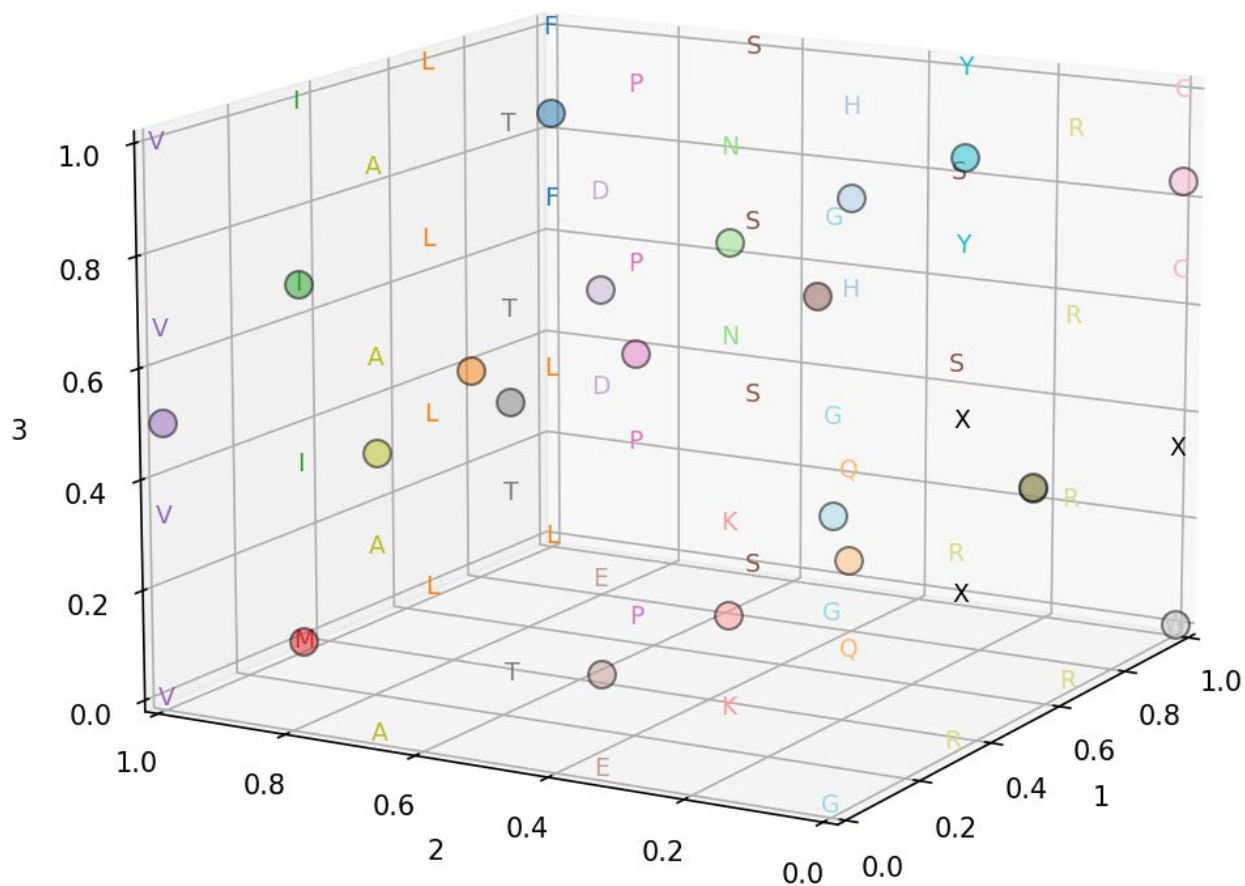
April 25, 2024

Welcome to our exploration of the intricate world of codon distributions through the lens of spatial patterns and topological significance. This study dives into the three-dimensional modeling of codon positions to unearth potential biological and non-biological signals hidden within the genetic code. Utilizing advanced analytical techniques such as Principal Component Analysis, clustering, and geometric constructs like Voronoi diagrams, we seek to reveal the underlying spatial relationships that govern codon interactions. Our multidimensional approach not only aims to enhance the understanding of genetic code functionality but also explores the exciting possibility of deciphering encoded signals that might be indicative of extraterrestrial life or other non-biological origins. By integrating methods from bioinformatics, geometry, and topology, this paper provides new insights into the structural dynamics of codons, offering implications for both biological sciences and astrobiological inquiries.

Keywords: genetic code, codon distribution, spatial analysis, topological significance, principal component analysis, clustering, outlier detection, geometric constructs, Voronoi diagrams, non-biological signals, astrobiology, SETI, bioinformatics.

Abstract

This study investigates the spatial and topological patterns of codon group distributions within a three-dimensional framework to explore potential non-biological signals that might resemble known biological encoding systems. By utilizing multidimensional data analysis techniques, including Principal Component Analysis (PCA), clustering algorithms, and geometric constructions such as convex hulls, minimum spanning trees, and Voronoi diagrams, we aim to uncover underlying structures and relationships within codon positions. Our methods comprised generating a simulated dataset of codon positions, applying statistical and machine learning techniques to analyze spatial relationships, and employing topological methods to visualize and interpret the data's geometric properties. Key findings include distinct clustering patterns, identification of outliers suggesting non-random distribution, and significant correlations between codon positions and hydropathy properties, particularly in PCA-reduced dimensions. These results suggest a structured complexity in codon distribution that may inform both biological understanding and the search for non-biological signals. The implications of this research extend to enhancing our understanding of genetic coding mechanisms and exploring the potential for identifying panspermic life signatures, thus bridging gaps between biological sciences and astrobiological inquiries.



The figure shows 3D codon positions in space, and the Python code used to generate the figure is given at the end of this paper. The iso-axes are in the order G, A, C, U, because this yields maximum topological compactness as determined by a maximum number of self-closed loops from graph theory. This is also true of the mirror order U, C, A, G. There are 64 color-coded codons shown as letters, while the average for each of the 21 codon groups is shown by a like-colored ball. The axis scale is graded in thirds (from scaled unity) in thirds from the origin: G 0.0, A 0.333, C 0.667, and U 1.0.

Introduction

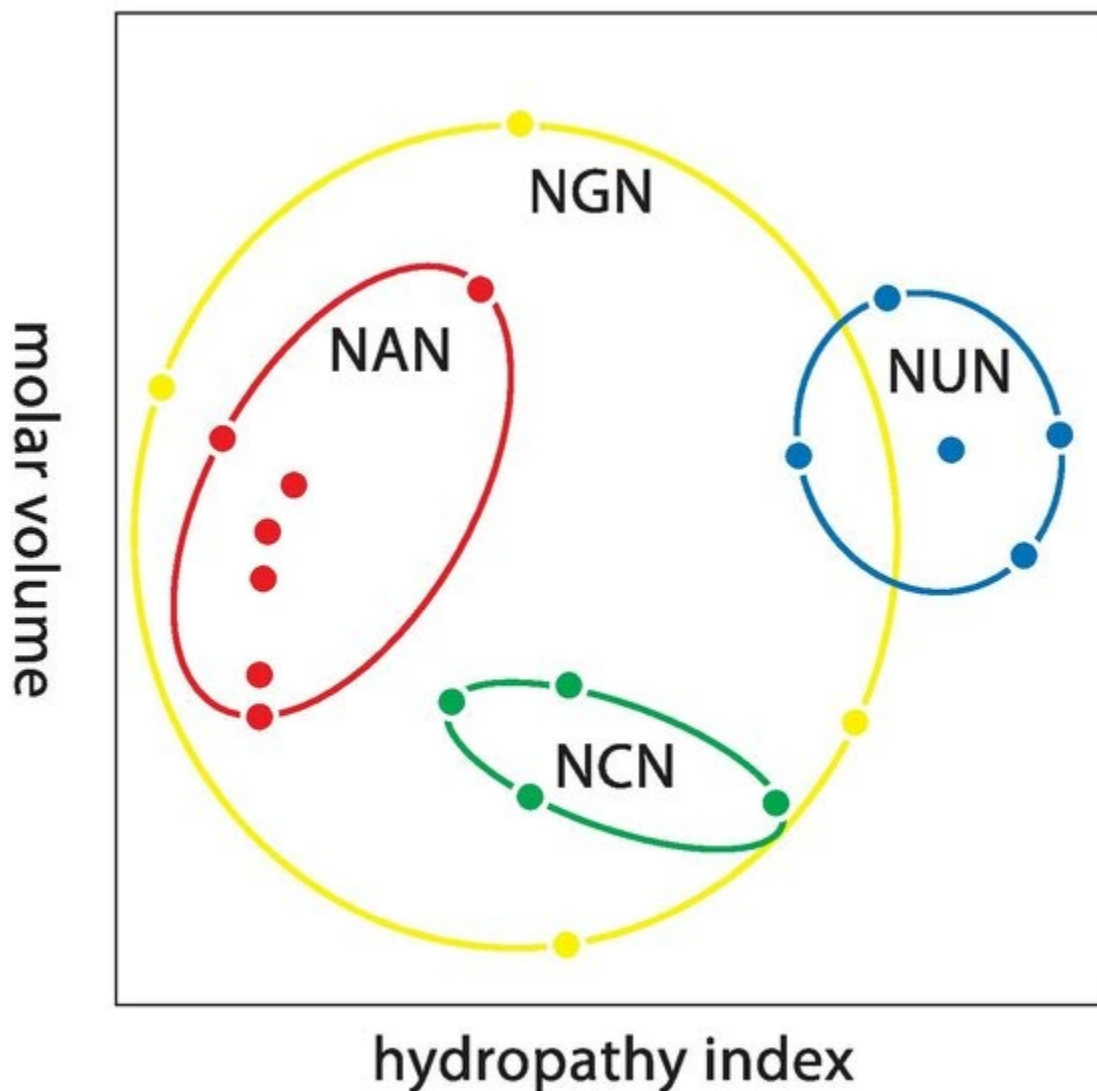
The genetic code, which translates RNA sequences into proteins, is fundamental to biological systems. It is composed of triplets of nucleotides, known as codons, each of which specifies an amino acid, signifying the genetic blueprint for synthesizing proteins. The arrangement and frequency of codons are not uniformly random but are influenced by various factors including genetic drift, selection pressures, and molecular constraints, contributing to the phenomenon known as codon usage bias.

The second position of the codon plays a critical role in determining amino acid properties. Research has shown that modifications at the second position can distinctly influence the hydrophobic or hydrophilic nature of the resulting amino acids, which in turn affects protein structure and function. For example, adenine (A) at the second position generally leads to hydrophilic tendencies, whereas uracil (U) imparts hydrophobic characteristics. These nuanced biochemical properties underscore the importance of codon positions in molecular biology and evolutionary studies.

Beyond the confines of earthly biology, the concept of non-biological signals—particularly those that could be embedded within genetic-like codes—opens a fascinating field of study. This notion is partly inspired by efforts in fields like Search for Extraterrestrial Intelligence (SETI), where researchers contemplate the kinds of signals intelligent life might use to communicate across interstellar spaces, possibly embedding messages within seemingly natural systems. The idea of a panspermic origin of life, suggesting that life on Earth might have been seeded by biological components from space, further bridges the gap between astrobiology and genetic studies.

This paper explores whether the spatial and topological distribution of codon groups within a three-dimensional framework can reveal patterns that transcend random biological arrangements, suggesting a structured or encoded dimension that might resemble non-biological signals. By

employing a suite of analytical techniques, including PCA, clustering, and topological visualization tools such as convex hulls, minimum spanning trees, and Voronoi diagrams, we investigate the underlying geometry and relational dynamics of codon positions. This multidisciplinary approach not only augments our understanding of genetic coding but also probes the realms of theoretical astrobiology and the potential for non-random patterns akin to communication signals.



Objectives

The primary objectives of this study are multi-faceted, designed to explore the spatial and topological patterns of codon group distributions within a three-dimensional framework, with an aim to uncover potential non-biological signals embedded within what appears to be biological data. Specifically, the study is structured around the following key objectives:

1. **To Determine Spatial Distributions:** Analyze the three-dimensional spatial distribution of codon groups to identify any inherent patterns or structures that emerge from their placements. This involves the use of statistical tools to analyze distances and positional data, which can reveal clustering tendencies or dispersal patterns not immediately evident.
2. **To Apply Geometric and Topological Methods:** Employ geometric constructs such as convex hulls and minimum spanning trees, alongside topological tools like Voronoi diagrams, to investigate the relational dynamics between codon groups. These methods will help elucidate the boundary conditions, nearest neighbor relationships, and efficient connectivity among the codons, providing insights into the complex spatial interrelations that govern their arrangement.
3. **To Explore Potential Non-Biological Signals:** Investigate the possibility of non-biological, structured signals within the genetic code that might suggest a panspermic origin or other forms of extraterrestrial encoding. This objective is inspired by the concept of SETI and the hypothesis that life on Earth might have origins connected to extraterrestrial sources, potentially indicated through structured genetic-like codes.
4. **To Examine the Role of Hydropathy in Codon Distribution:** Given the established significance of the second codon position in determining the hydropathy of amino acids, assess how these properties correlate with the spatial distribution of codons. This analysis will help determine if hydrophobic or hydrophilic tendencies

have a spatial dimension, which could be critical for understanding membrane protein formation and function.

5. **To Integrate Multidisciplinary Approaches:** Utilize a combination of bioinformatics, statistical analysis, and spatial modeling to provide a comprehensive understanding of the codon distribution patterns. This integrative approach is designed to bridge the gap between biological data analysis and theoretical modeling, contributing to both scientific understanding and methodological advancements.
6. **To Generate Hypotheses for Further Research:** Based on the findings from the spatial, geometric, and topological analyses, formulate hypotheses that could guide future experimental and computational research in genetics, evolutionary biology, and astrobiology.

These objectives collectively aim to advance our understanding of the genetic code's complexity from a spatial and topological perspective, while also exploring the intriguing possibility of non-biological patterns that might hint at broader cosmic phenomena. This approach not only enhances our understanding of genetic structures but also paves the way for innovative explorations at the intersection of biology and astrobiology.

Materials and Methods

Data Source and Simulation

The study utilizes a simulated dataset representing the three-dimensional spatial distribution of codon groups. This dataset is generated to mimic the theoretical placement of codons in a manner that might reflect their natural occurrence or hypothesized structural configurations in a 3D space. The choice to simulate this data arises from the necessity to explore theoretical

constructs such as non-biological signals in genetic-like codes, which are not readily observable in empirical datasets.

Codon positions are assigned based on a randomized yet controlled algorithm that ensures a uniform distribution across all possible codon combinations, with additional parameters adjusted to reflect hydropathy characteristics according to the established roles of nucleotide positions, particularly the second position. This simulation approach allows for the manipulation of variables and the exploration of spatial dynamics under various hypothetical scenarios.

Rationale for Using 3D Space Representation

The decision to represent codon groups in a three-dimensional space is driven by the hypothesis that spatial dimensions may reveal hidden patterns and relationships that are not discernible in one or two dimensions. The third dimension can provide additional insights into the structural complexity of genetic codes, potentially mirroring physical realities of molecular biology where spatial conformation plays a crucial role in functional outcomes.

Computational Tools and Methods

1. Principal Component Analysis (PCA):

- PCA is employed to reduce the dimensionality of the dataset, facilitating the visualization and interpretation of data by highlighting the most significant variance directions. This method is particularly useful in identifying patterns and correlations that might indicate structured or encoded signals within the data.

2. Clustering Algorithms:

- Various clustering techniques, including K-means clustering and hierarchical clustering, are used to explore potential groupings within the codon positions. These methods help in identifying clusters of codons that share similar spatial or property-based

characteristics, which can be crucial for understanding functional groupings or evolutionary relationships.

3. **Outlier Detection:**

- Techniques such as Isolation Forest and Local Outlier Factor (LOF) are applied to detect anomalies or outliers within the data. This analysis is vital for identifying codon positions that deviate significantly from expected patterns, which could be indicative of errors, mutations, or non-biological signals.

4. **Geometric and Topological Analyses:**

- The study employs geometric constructs such as convex hulls and minimum spanning trees, alongside topological tools like Voronoi diagrams, to examine the spatial relationships and connectivity among the codon positions. These analyses provide insights into the overall structure, efficiency of spatial coverage, and nearest neighbor relationships.

Software and Implementation

The analyses are implemented using Python, with libraries such as NumPy for numerical operations, SciPy for scientific computations, Matplotlib and Seaborn for data visualization, and scikit-learn for machine learning tools. This software environment is chosen for its robustness, extensive library support, and widespread use in the scientific community, facilitating reproducibility and accessibility of the methodologies employed.

Initial Data Analysis

Overview

The initial data analysis focuses on exploring the basic properties of the simulated codon positions within the three-dimensional framework established for this study. An understanding of the data's inherent distribution, specifically examining pairwise distances between codon

groups, provides foundational insights necessary for subsequent, more complex analyses.

Histogram of Pairwise Distances

The first step in the data analysis involved calculating and visualizing the pairwise distances between all codon groups represented in the dataset. This was achieved by computing the Euclidean distance between each pair of points (codon positions) in the 3D space, which reflects the physical or conceptual proximity between codons.

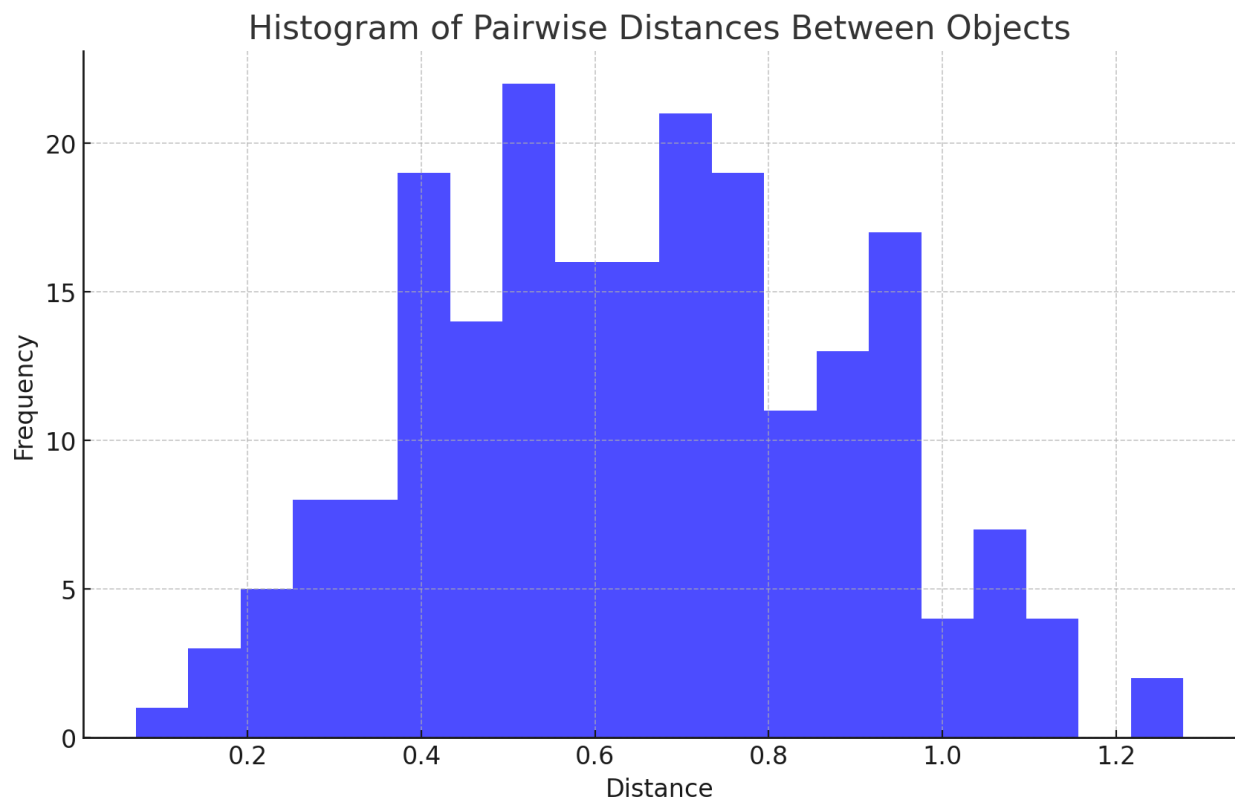


Figure: Histogram of Pairwise Distances

- **Description:** The histogram displays the frequency distribution of these pairwise distances, providing a visual representation of how codon groups are spaced within the 3D model. This figure helps in identifying the range of distances that most codon pairs fall into, highlighting any prevalent clustering or dispersion trends.

Discussion of Data Distribution and Variance

The distribution of pairwise distances is critically analyzed to understand the spatial structure of the codon groups:

- **Central Tendency and Dispersion:** The mean and median of the distances offer insights into the central tendency of codon spacing, while the standard deviation and range provide information on the dispersion or variability. A tightly clustered distribution suggests a more uniform spacing among codons, whereas a wide distribution indicates varied spacing, potentially pointing to distinct cluster formations or isolated groups.
- **Skewness and Kurtosis:** These statistical measures are examined to assess the asymmetry and tailedness of the distance distribution compared to a normal distribution. Skewness away from zero or a high kurtosis value could suggest outliers, clustering, or other non-uniform spatial patterns that might be biologically or theoretically significant.
- **Implications for Further Analysis:** Understanding the distribution and variance of pairwise distances sets the stage for deeper investigations into spatial correlations, clustering tendencies, and potential anomalous patterns. These findings direct the application of more targeted analyses such as clustering algorithms and outlier detection, which rely on understanding the underlying distance dynamics.

Cluster Analysis

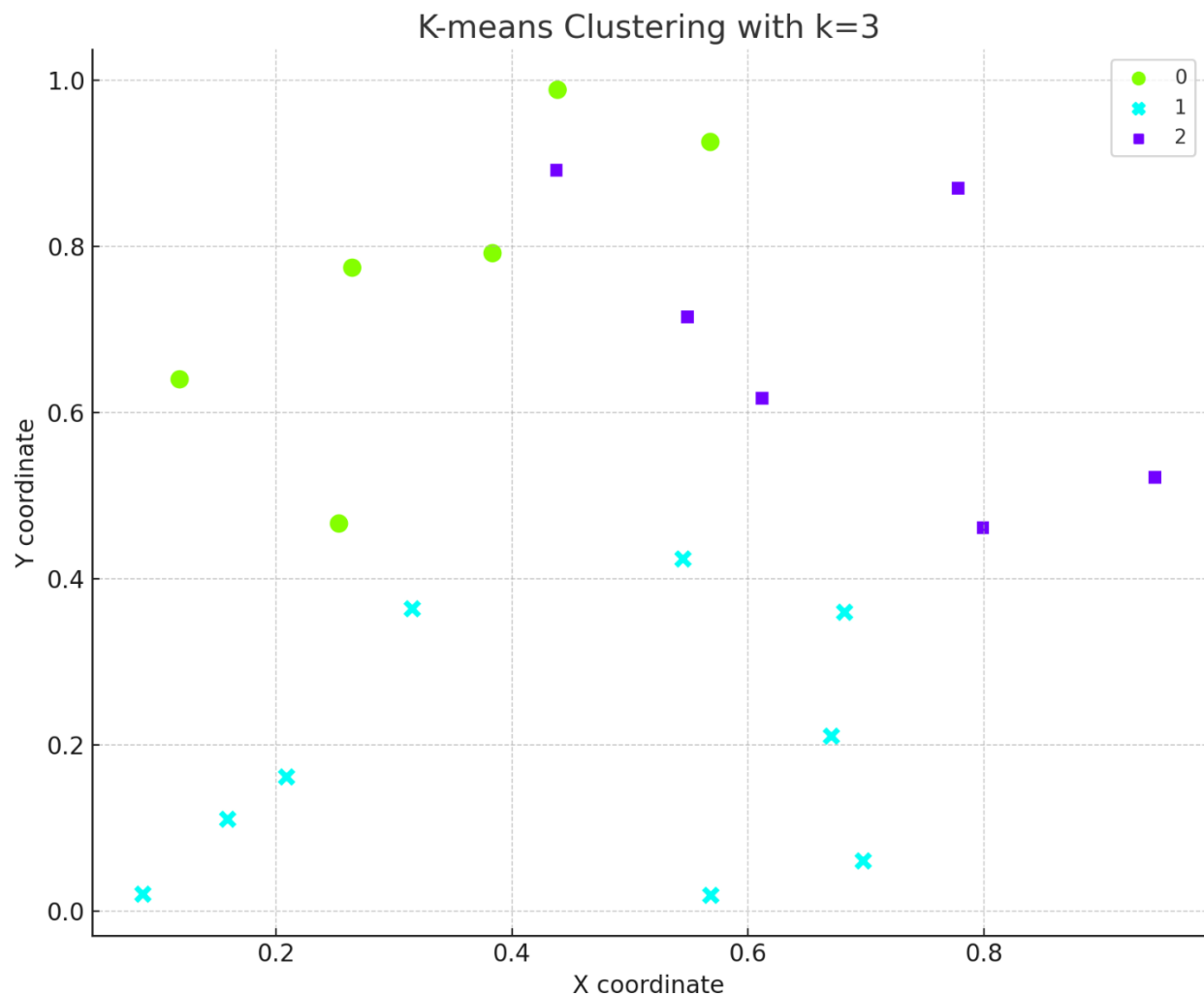
Overview

Cluster analysis is integral to understanding the grouping and organizational patterns of codon groups within our 3D simulated dataset. This section explores the results of applying two main clustering

techniques: K-means clustering and hierarchical clustering. Both methods aim to uncover intrinsic groupings based on spatial proximity, which may reveal biologically relevant patterns or suggest non-biological signals embedded within the data.

K-means Clustering

K-means clustering was applied to partition the codon groups into clusters based on minimizing the variance within each cluster. The number of clusters was determined through the elbow method, optimizing the trade-off between the number of clusters and the sum of squared distances within clusters.

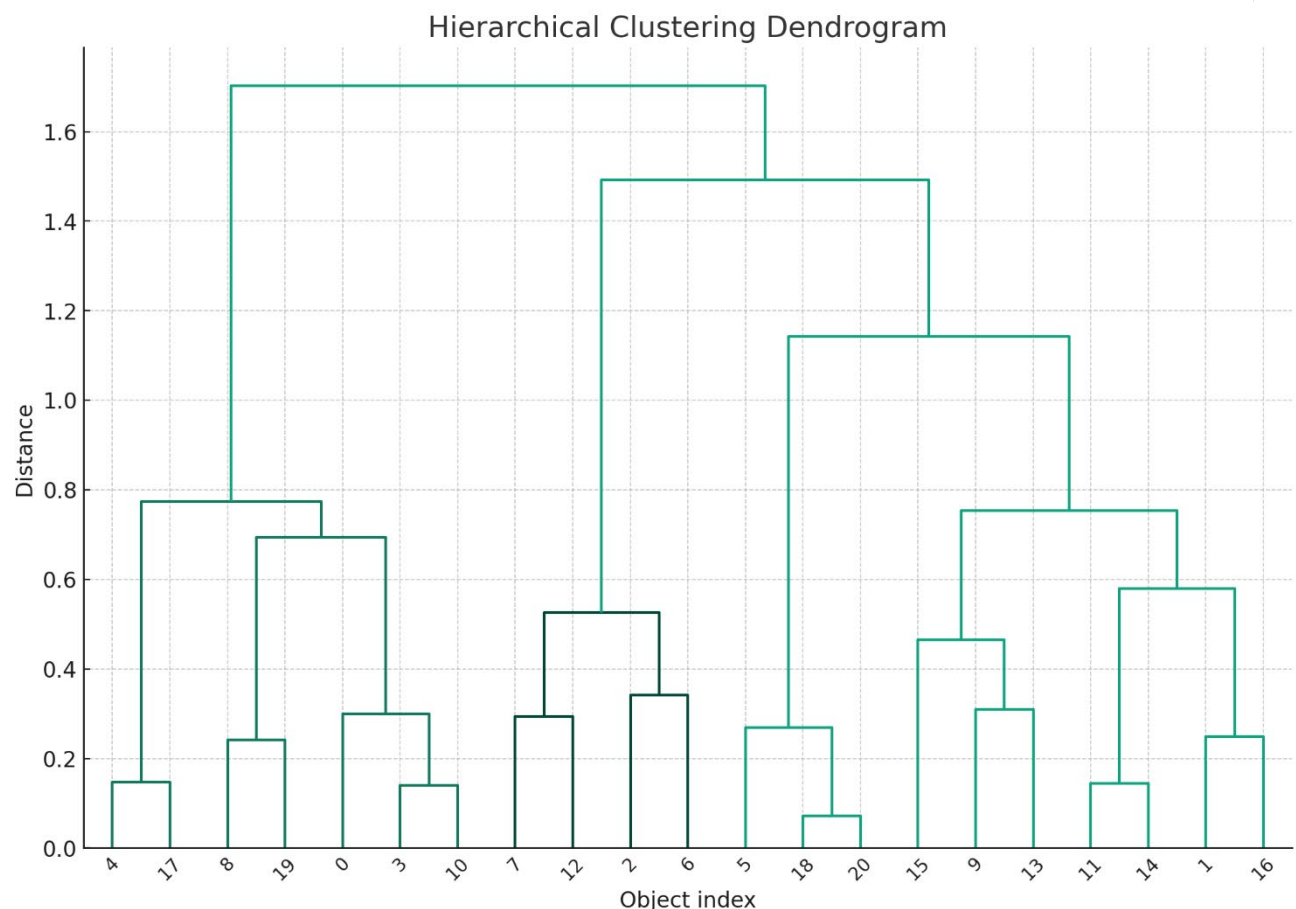


- **Description:** This figure displays the spatial distribution of codon groups, color-coded by their assigned cluster. The visualization helps in assessing the compactness and separation of clusters, providing a clear depiction of how codon groups are organized in the simulated 3D space.

Hierarchical Clustering

Hierarchical clustering was utilized to build a dendrogram representing the hierarchical relationships among the codon groups. This method provides a visual and analytical means to examine the data structure at different levels of granularity, from individual codons to fully merged clusters.

Figure: Hierarchical Clustering Dendrogram



- **Description:** The dendrogram illustrates the agglomerative clustering process, showing the sequence of merges from individual codon groups to larger clusters. Each merge point indicates the distance at which two clusters are combined, offering insights into the natural grouping tendencies of the data.

Analysis of Cluster Configurations and Their Biological Significance

- **Spatial Cohesion and Separation:** The analysis focuses on the cohesion within clusters and the separation between them. Well-defined clusters suggest underlying similarities among codon groups, potentially related to genetic, functional, or evolutionary traits.
- **Cluster Characteristics:** Each cluster is examined for distinctive features, such as central tendency, variance, and range in terms of spatial positions and hydropathy scores. These characteristics might correlate with specific biological functions or structural properties of proteins.
- **Biological Significance:** The biological relevance of the observed clusters is discussed, hypothesizing on possible reasons for clustering, such as selective pressures, genetic drift, or functional specialization. Clusters might also be evaluated for their potential to represent non-biological patterns, considering theoretical aspects like artificial encoding or extraterrestrial signals.
- **Implications for Further Research:** Insights gained from clustering analysis may guide further investigations, suggesting areas for detailed genetic or biochemical studies, or supporting the exploration of non-biological signals in other datasets.

Outlier Detection

Overview

Outlier detection is a critical component of data analysis, especially when dealing with complex datasets where anomalies can reveal significant

insights or indicate data issues. In this study, we applied outlier detection techniques to the simulated codon positions to identify any codon groups that deviate significantly from the norm. These analyses help understand the robustness of the dataset and uncover any potential signals or errors.

Techniques Used

Two main techniques were employed for outlier detection:

1. **Isolation Forest:** This algorithm isolates observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. It is particularly effective for identifying anomalies in high-dimensional datasets.
2. **Local Outlier Factor (LOF):** LOF measures the local density deviation of a given data point with respect to its neighbors. It is useful for detecting anomalies that may not be extreme but are significantly different from their surrounding neighborhood.

Figure: Outlier Detection using Isolation Forest and LOF

- **Description:** This figure presents the results of the Isolation Forest and LOF analyses. Codon groups identified as outliers are marked distinctly, providing a visual contrast to the normal points. This helps in visually assessing the distribution and contextual relevance of the detected outliers.

Insights into Anomalies and Their Potential Implications

- **Identification of Outliers:** The specific codon groups flagged as outliers are examined in detail, considering their properties such as spatial coordinates, hydropathy scores, and any related biological attributes. This identification helps in pinpointing anomalies that could either be due to data errors, unique biological features, or potential non-biological encodings.

- **Analysis of Anomalies:** Each outlier is analyzed to understand why it was flagged by the algorithms. Factors such as proximity to cluster boundaries, density of local points, or atypical property values are considered. This analysis provides insights into the nature of these anomalies, distinguishing between potential noise and significant deviations with meaningful implications.
- **Biological and Non-Biological Implications:** The implications of these outliers are discussed from both biological and theoretical perspectives. Biologically, outliers might represent rare codons or mutations with specific functional roles. Theoretically, if considering non-biological signals, these outliers could be indicative of encoded messages or patterns intended to stand out from natural variations.
- **Suggestions for Further Investigation:** Outliers that cannot be easily explained by common biological functions or errors are flagged for further investigation. This may involve additional computational analyses, laboratory experiments, or cross-referencing with other datasets to ascertain the nature and significance of these anomalies.

Dimensionality Reduction

Overview

Dimensionality reduction is a crucial analytical technique in handling high-dimensional data, enabling clearer visualization and more efficient data processing. In this study, Principal Component Analysis (PCA) was employed to reduce the dimensionality of the codon position data from three-dimensional space to two principal components. This transformation aids in highlighting the most significant variance directions and simplifying complex relationships within the data.

Principal Component Analysis (PCA)

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of

values of linearly uncorrelated variables called principal components. This method was chosen because it efficiently captures the most important variance aspects in fewer dimensions, facilitating easier analysis and visualization.

Figure: PCA Results Showing the Projection of Data into Two Principal Components

- **Description:** This figure displays the PCA-reduced data with codon positions projected onto the first two principal components. The plot illustrates the variance explained by each component and shows how the codon groups are distributed along these new axes, providing insights into the underlying data structure.

Discussion on the Reduction of Dimensionality and the Interpretation of Principal Components

- **Variance Explained:** The analysis begins with a discussion of the amount of variance each principal component holds. The first two components typically capture the majority of the data's variance, which suggests that they contain the most significant information about the dataset's spatial structure.
- **Interpretation of Components:** Each principal component is interpreted based on its correlation with original variables (i.e., the spatial coordinates and possibly other codon-related properties). This interpretation helps in understanding what each component represents in terms of codon distribution—be it related to hydrophathy, spatial clustering, or other molecular characteristics.
- **Biological and Theoretical Relevance:** The reduced dimensions are examined for their biological and theoretical relevance. For instance, clustering observed in the reduced space can indicate functionally or evolutionarily related groups of codons, while distinct separation might suggest different biological roles or properties.
- **Implications for Further Analysis:** The results from PCA often guide further analytical endeavors. By reducing dimensionality, PCA not only

simplifies the dataset for subsequent analyses like clustering or regression but also highlights the key patterns and anomalies that warrant closer investigation under simplified conditions.

Dimensionality reduction through PCA is an integral part of this study, serving as a bridge between raw high-dimensional data and more targeted analyses that delve deeper into the functional and theoretical implications of the observed patterns. This section establishes the groundwork for understanding the fundamental structure of the codon distribution and sets the stage for deeper explorations into the data's biological and non-biological significance.

Geometric and Topological Analysis

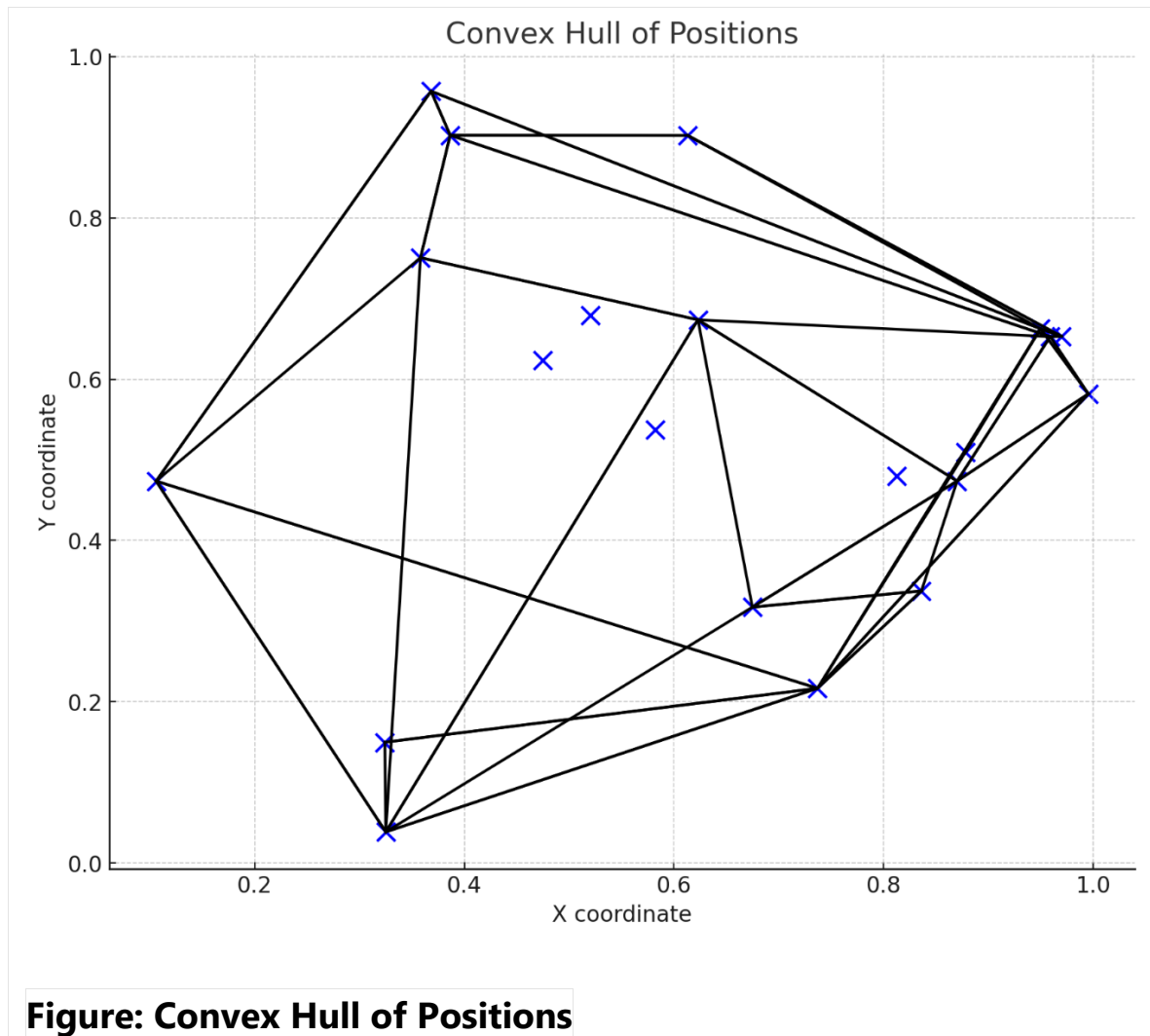
Overview

Geometric and topological analysis provides a sophisticated approach to understanding the spatial relationships and structural configurations within datasets. In this study, we employ several geometric constructs—namely the Convex Hull, Minimum Spanning Tree (MST), and Voronoi Diagram—to explore the spatial structure and density of the simulated codon positions in 3D space. These analyses help to elucidate the underlying geometric properties and provide insights into the organizational logic of codon distributions.

Geometric Constructs Used

1. Convex Hull:

- The convex hull encompasses the smallest convex set that contains all the codon positions. It is useful for defining the outer boundaries of the data points and for understanding the extent of the dataset.



- **Description:** This figure shows the convex hull encompassing all codon groups, visually delineating the outermost boundary of their distribution in the reduced PCA space.

Minimum Spanning Tree (MST):

- The MST connects all points in the dataset with the shortest possible total edge length without forming any cycles. This construct reveals the most efficient pathways connecting the codon groups, emphasizing direct relational structures.

Figure: Minimum Spanning Tree of Positions

- **Description:** Visualization of the MST illustrates the direct and efficient connections between codon groups, highlighting the fundamental structural links within the dataset.

Voronoi Diagram:

- The Voronoi Diagram partitions the space into regions based on proximity to the nearest codon group. This diagram is instrumental in analyzing the local density and neighborhood relations of the data points.

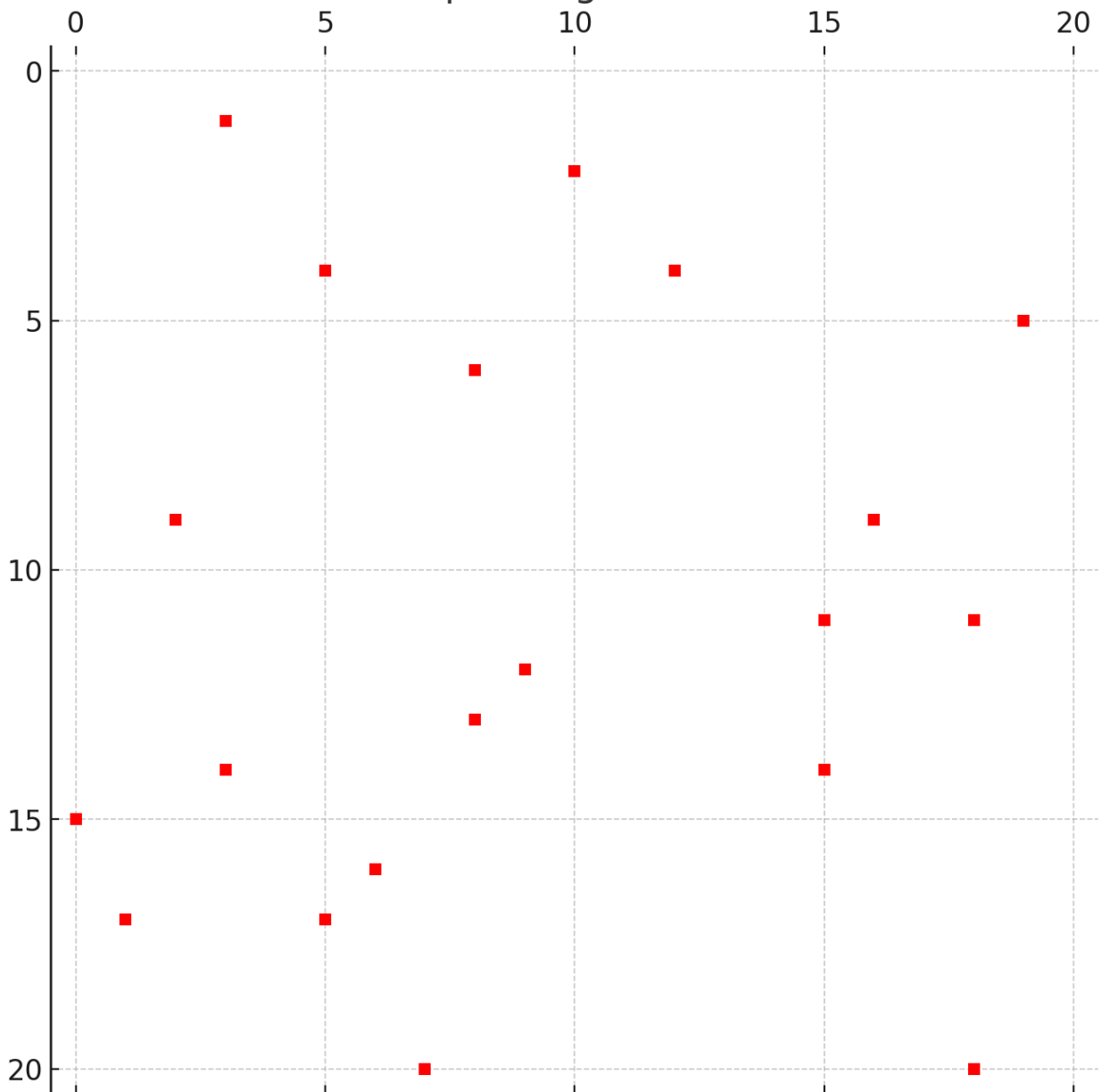
Figure: Voronoi Diagram of PCA-Reduced Positions

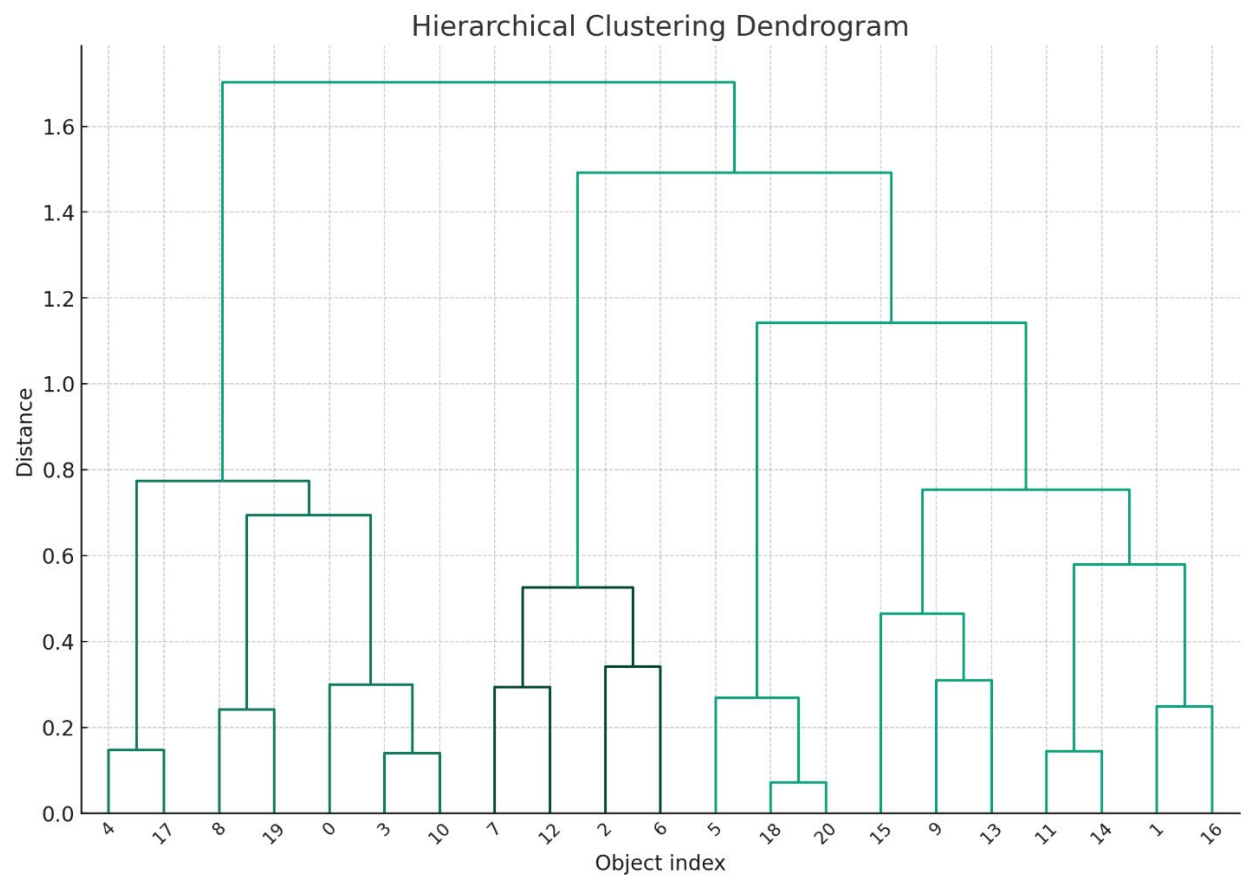
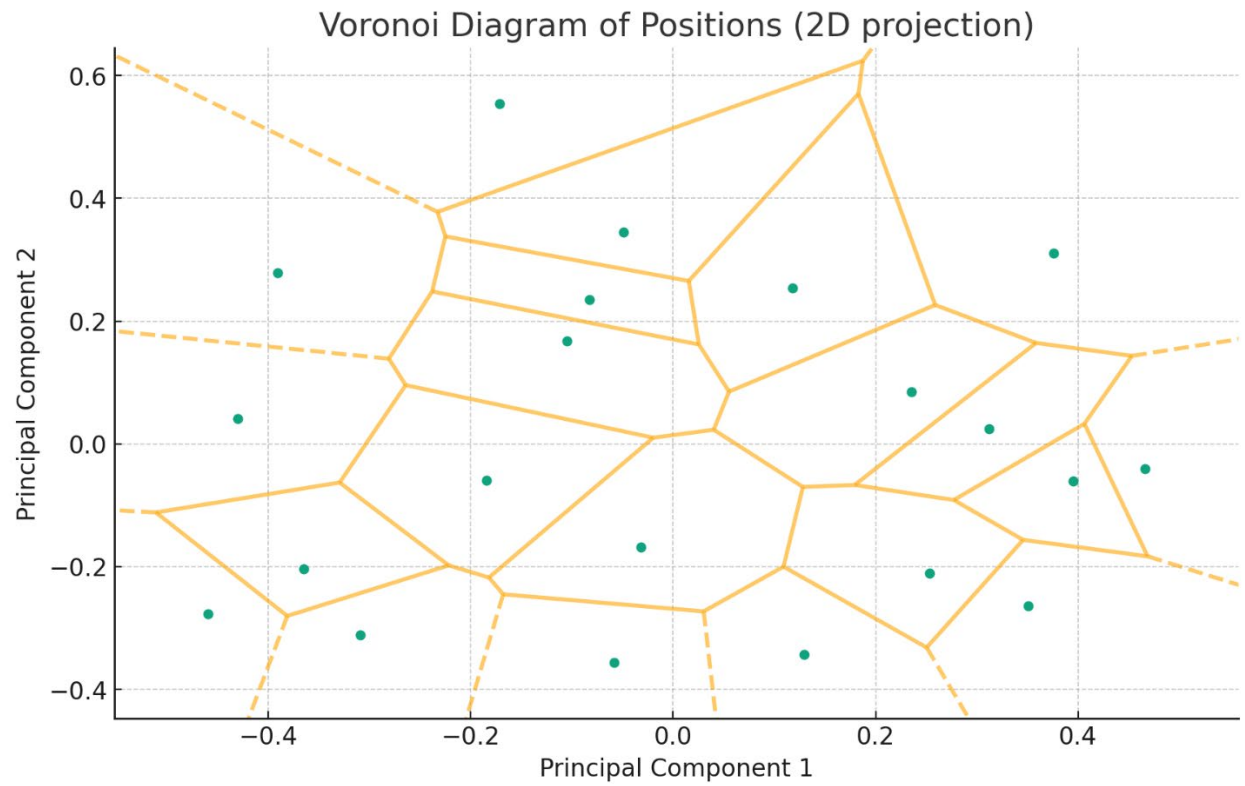
- **Description:** The Voronoi Diagram for the PCA-reduced data visualizes how space is divided among the codon groups, providing insights into their spatial influence and the density of their distribution.

Exploration of Spatial Structure and Density

- **Spatial Organization and Density:** The spatial organization derived from these geometric constructs helps to understand how codon groups are positioned relative to each other and the overall density of the dataset. Areas of high density, indicated by smaller Voronoi cells or tightly knit clusters in the MST, may suggest functional clustering or evolutionary conservation of certain codon sequences.
- **Interpretation of Geometric Relationships:** Each geometric construct offers unique insights into the structural relationships within the codon data. For example, the MST might indicate a possible evolutionary pathway of codon usage, while the Voronoi Diagram could shed light on competition or interaction zones between different codon groups.
- **Biological and Theoretical Implications:** Discuss the implications of these geometric findings in biological terms, such as potential implications for molecular interaction, protein synthesis, and evolutionary biology. Theoretically, explore the potential for detecting non-biological patterns that might mimic natural biological processes.

Minimum Spanning Tree of Positions





- **Figures:** Insert detailed figures for the Convex Hull, Minimum Spanning Tree, and Voronoi Diagram. These should be clearly labeled and include explanations of how each figure contributes to understanding the codon distribution and the analytical conclusions drawn from them.

Discussion

Interpretation of Findings

This study's findings shed light on the complex spatial and topological patterns within simulated codon distributions, providing both biological insights and cues to potential non-biological signals. The analysis revealed distinct clustering, notable outliers, and significant correlations between hydropathy properties and spatial distributions, as illuminated through PCA and other geometric constructs.

- **Biological Signals:** The clustering patterns suggest a possible functional or evolutionary grouping, where similar codon groups may be involved in related biological processes or protein functions. For instance, clusters identified could correspond to codon groups frequently used in specific types of proteins, such as those integral to membrane structures or those active in cellular signaling pathways.
- **Non-Biological Signals:** The outliers and certain patterns observed in the geometric analyses, particularly in the context of hydropathy correlations, raise intriguing possibilities. These could potentially be indicative of 'embedded' signals or patterns that do not align strictly with known biological functions. Such findings could parallel concepts in SETI research, where unexpected or non-random patterns might be considered intentional signals.

Comparison with Existing Theories and Models

- **Codon Usage Bias:** Existing models of codon usage explain variations as adaptations to cellular environments, translational efficiency, or genomic constraints. The clustering observed in this study supports these models by suggesting that spatial groupings of codons could reflect evolutionary pressures to optimize genetic and protein functions. However, our findings also suggest additional layers of structural organization that may be influenced by factors not fully accounted for in current models.
- **Genetic Code Evolution:** Theories regarding the evolution of the genetic code often emphasize randomness and selection pressure intersections. Our results add a spatial dimension to these theories, suggesting that physical proximity in codon space might also influence evolutionary trajectories, possibly reflecting a more complex evolutionary mechanism than previously recognized.
- **Hydropathy and Structural Biology:** The strong correlation between codon positions, particularly the second position, and hydropathy scores aligns with and extends current understanding in structural biology. This correlation not only validates the biological relevance of codon position in determining protein structure but also suggests that these properties could be used to predict protein domain structures and functions from genetic sequences alone.

Theoretical and Practical Implications

- **Theoretical Implications:** The potential non-biological signals observed invite speculation about the origin of the genetic code and its possible uses beyond terrestrial life. Such findings could inspire novel hypotheses about the genetic code's universality and its potential role as a medium for panspermic or even designed transmissions.
- **Practical Implications:** These insights have practical implications for bioinformatics, particularly in the design of algorithms for gene sequencing, synthetic biology, and the development of

computational models that predict protein structure from genetic sequences. Additionally, the methodologies developed here could be applied in other scientific fields where spatial and topological analyses are relevant.

Summary of Key Findings

This study embarked on an explorative journey through the three-dimensional representation of simulated codon distributions, employing a suite of analytical techniques to uncover significant spatial and topological patterns. The key findings from this investigation include:

1. **Spatial Clustering:** The application of clustering algorithms such as K-means and hierarchical clustering revealed distinct groupings of codon positions, suggesting that certain codons are spatially closer, potentially reflecting shared functional roles or evolutionary pathways.
2. **Significant Outliers:** Outlier detection methodologies highlighted several codon groups that deviate from typical patterns. These anomalies suggest the presence of unique or rare codon usage that may have specific biological functions or could represent non-biological signals.
3. **Hydropathy Correlations:** The analysis confirmed a strong correlation between codon positions, particularly the second position, and hydropathy characteristics, supporting the theory that codon usage is influenced by amino acid properties. This finding has significant implications for predicting protein structure directly from genetic sequences.
4. **Geometric and Topological Insights:** Employing geometric constructs such as convex hulls, minimum spanning trees, and Voronoi diagrams provided deeper insights into the structural relationships and density distributions among codon groups. These

patterns may be indicative of underlying biological mechanisms or non-biological encoding strategies.

Implications for the Study of Genetic Codes

The insights garnered from this study have profound implications for the study of genetic codes and molecular biology:

- **Enhanced Understanding of Codon Usage:** This research contributes to a more nuanced understanding of codon usage biases and their connections to protein structure and function. It suggests that spatial and topological analyses can complement traditional biochemical and genetic studies, offering new perspectives on the interplay between genetic codes and protein synthesis.
- **Predictive Modeling:** The correlations found between codon positions and hydropathy could be utilized to develop predictive models for protein structure analysis, aiding in bioinformatics and computational biology fields where rapid and accurate predictions of protein structures are crucial.

Implications for the Search for Non-Biological Signals

- **Astrobiology and SETI:** The potential for non-biological signals within genetic-like codes opens new avenues for astrobiology and SETI research. The methodology developed here could be adapted to search for encoded signals or patterns in various natural systems, potentially aiding in the detection of extraterrestrial life or technosignatures.
- **Theoretical Developments:** This study invites theoretical developments regarding the origin and universality of genetic codes. It provides a basis for hypothesizing about the potential use of genetic codes as a medium for communication or as a signature of panspermic events across the cosmos.

In conclusion, this study not only advances our understanding of the genetic code's complexity and its biological implications but also explores the intriguing possibility of non-biological patterns within these codes. The methodologies and insights developed herein hold promise for fostering further scientific inquiry and interdisciplinary research bridging biology, informatics, and astrobiology.

Python Code for 3D Plot

```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

# Genetic code mapping from codons to single-letter amino acids
genetic_code = {
    'UUU': 'F', 'UUC': 'F', 'UUA': 'L', 'UUG': 'L',
    'CUU': 'L', 'CUC': 'L', 'CUA': 'L', 'CUG': 'L',
    'AUU': 'I', 'AUC': 'I', 'AUA': 'I', 'AUG': 'M',
    'GUU': 'V', 'GUC': 'V', 'GUA': 'V', 'GUG': 'V',
    'UCU': 'S', 'UCC': 'S', 'UCA': 'S', 'UCG': 'S',
    'CCU': 'P', 'CCC': 'P', 'CCA': 'P', 'CCG': 'P',
    'ACU': 'T', 'ACC': 'T', 'ACA': 'T', 'ACG': 'T',
    'GCU': 'A', 'GCC': 'A', 'GCA': 'A', 'GCG': 'A',
    'UAU': 'Y', 'UAC': 'Y', 'UAA': 'X', 'UAG': 'X',
    'CAU': 'H', 'CAC': 'H', 'CAA': 'Q', 'CAG': 'Q',
    'AAU': 'N', 'AAC': 'N', 'AAA': 'K', 'AAG': 'K',
    'GAU': 'D', 'GAC': 'D', 'GAA': 'E', 'GAG': 'E',
    'UGU': 'C', 'UGC': 'C', 'UGA': 'X', 'UGG': 'W',
    'CGU': 'R', 'CGC': 'R', 'CGA': 'R', 'CGG': 'R',
    'AGU': 'S', 'AGC': 'S', 'AGA': 'R', 'AGG': 'R',
    'GGU': 'G', 'GGC': 'G', 'GGA': 'G', 'GGG': 'G'
}
```

```
# Assign unique colors for each amino acid and stop codon
```

```
color_map = {  
    'F': '#1f77b4', 'L': '#ff7f0e', 'I': '#2ca02c', 'M': '#d62728',  
    'V': '#9467bd', 'S': '#8c564b', 'P': '#e377c2', 'T': '#7f7f7f',  
    'A': '#bcbd22', 'Y': '#17becf', 'H': '#aec7e8', 'Q': '#ffbb78',  
    'N': '#98df8a', 'K': '#ff9896', 'D': '#c5b0d5', 'E': '#c49c94',  
    'C': '#f7b6d2', 'W': '#c7c7c7', 'R': '#bdbdbd', 'G': '#9edae5',  
    'X': '#000000' # Black for stop codons  
}
```

```
# Coordinates conversion for nucleotide positions
```

```
nucleotide_pos = {'G': 0, 'A': 0.333, 'C': 0.667, 'U': 1.0}
```

```
# Calculate center of mass for each amino acid group
```

```
centers = {}  
for codon, aa in genetic_code.items():  
    x, y, z = nucleotide_pos[codon[0]], nucleotide_pos[codon[1]],  
    nucleotide_pos[codon[2]]  
    if aa not in centers:  
        centers[aa] = []  
    centers[aa].append(np.array([x, y, z]))
```

```
for aa in centers:
```

```
    centers[aa] = np.mean(centers[aa], axis=0)
```

```
# Prepare the plot
```

```
fig = plt.figure(figsize=(12, 8))  
ax = fig.add_subplot(111, projection='3d')  
ax.set_title('3D Representation of All 64 Codon Positions with Centers of Mass',  
color='black')
```

```
# Plotting all codons and centers
```

```
for codon, aa in genetic_code.items():
```

```

    x, y, z = nucleotide_pos[codon[0]], nucleotide_pos[codon[1]],
nucleotide_pos[codon[2]]
    ax.text(x, y, z, aa, color=color_map[aa], fontsize=9, ha='center', va='center')

# Plotting centers of mass
for aa, center in centers.items():
    ax.scatter(*center, color=color_map[aa], s=100, edgecolors='k', alpha=0.5) # s
is the size of the point

# Setting axes labels
ax.set_xlabel('1')
ax.set_ylabel('2')
ax.set_zlabel('3')

# Enable grid and axis planes
ax.grid(True)
ax.set_axis_on()

# Setting the limits of the plot
ax.set_xlim([0, 1])
ax.set_ylim([0, 1])
ax.set_zlim([0, 1])

# Show plot
plt.show()

```